DOCUMENT RESUME

ED 153 646 IR 006 028

AUTHOR Misanchuk, Earl R.
TITLE Evaluation Types in Instructional Development: Beyond the Formative-Summative Distinction.
PUB DATE 78
NOTE 16p.; Paper presented to the Association for Educational Communications and Technology (Kansas City, Missouri, April 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Communication Problems; Decision Making; *Evaluation; *Formative Evaluation; *Instructional Design; *Program Evaluation; *Summative Evaluation; *Vocabulary

ABSTRACT

In the context of instructional development projects, situations arise in which evaluations are required that are neither formative nor summative in role. Typical responses to such situations are to stretch the meaning of one or the other of the terms introduced by Scriven (1967) to accommodate the situation; a consequence is a loss in precision of language which often results in difficulties in communication among practitioners. Another role descriptor, confirmative evaluation, describes a situation in which an operational program is brought up for review, and its characteristics are explored. Additional dimensions of evaluation, two developed by Davies and Schwen (1972), heuristic-algorithmic and determinate-indeterminate, provide detail regarding the procedures and data types used in the evaluation; a third, acquirement-accomplishment evaluation, describes the kind of outcome or behavior under scrutiny. Using the various dimensions in reporting evaluation studies could lead to more precise communication. (Author/VT)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED153646

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EVALUATION TYPES IN INSTRUCTIONAL DEVELOPMENT: BEYOND THE FORMATIVE-SUMMATIVE DISTINCTION

Earl R. Misanchuk
The University of Saskatchewan

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Earl R. Misanchuk

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Paper presented to the Annual Meeting,
Association for Educational Communications and Technology
Kansas City, Missouri, April 17-21, 1978

© E.R. Misanchuk, 1978

IR006028

ABSTRACT

In the context of instructional development projects, situations arise in which evaluations are required that are neither formative nor summative in role. Typical responses to such situations are to stretch the meaning of one or the other of the terms introduced by Scriven (1967) to accommodate the situation; a consequence is a loss in precision of language which often results in difficulties in communication among practitioners. Another role descriptor, confirmative evaluation, is introduced to describe a situation in which an operational program is brought up for review, and its characteristics are explored. Additional dimensions of evaluation are described: two dimensions developed by Davies and Schwen (1972) are recapitulated, and a third dimension, acquirement vs. accomplishment evaluation, is proposed. Using the various dimensions in reporting evaluation studies, it is argued, will lead to more precise communication.

As the field of Instructional Development (ID) emerges and grows, it is of fundamental importance that practitioners in the field be able to communicate completely and fully with one another--to share plans, experiences, and outcomes; to consolidate the resource base of knowledge regarding the efficacy of various practices. Instructional developers need not be reminded of the importance of full and complete communication, and the potential problems inherent therein--most of them deal with that very problem on a day-to-day basis.

I would like to focus on communication problems due to vocabulary in a particular subset of the terminology used by practitioners of instructional development, which is central to communicating about most ID projects: evaluation.

As we consider evaluation in the context of ID projects, I would ask you to recall some of the terms commonly used to describe different types and functions of evaluations.

Since their introduction by Scriven (1967), the terms "formative evaluation" and "summative evaluation" have so thouroughly permeated the field of education, and extended beyond into the fields of health and social action programs, that they have become commonplace. Along the way, their meanings have sometimes assumed different shades than those originally proposed by Scriven.

Let me illustrate. Some time ago, I had a visit from a young lady who was the chief executive officer of what was, in effect, a half-way house for teenage girls who had left home and were having difficulty adapting to and coping with their community environment. She sought my advice on how she should go about doing a "summative evaluation" on the three-year experimental program whose funding by the provincial government would be coming to an end in two months. Her board of directors, to whom she was personally responsible, wanted a high-quality evaluation, so she was shopping through the University for guidance. As we discussed the matter, I discovered that there had been no evaluation planned as part of the original

proposal for the development of the halfway house (the evaluation component was tacked on by the funding agency); that there had been no previous (formative) evaluation conducted as plans and operational procedures had been implemented; and that there was, according to her, absolutely no possibility that the funding agency or any similar agency would invest money in the project in the future. She admitted that the mode of operation of the home was unlikely to change in response to the findings of the evaluation, and that her primary motive for conducting the evaluation was to satisfy the requirement of the expiring funding contract. Furthermore, she said, there was virtually no hope that the program would be expanded or adopted elsewhere, largely because the informal structure of the program was undocumented, and the program, therefore, was not portable.

I suggest that what the lady was searching for cannot properly be called a summative evaluation. A summative evaluation, according to the definition given by Scriven, is an evaluation of a finished product, which would be most useful to someone considering the purchase or adoption of that product. A summative evaluation, by definition, describes what a (finished) object of evaluation is capable of doing. Obviously, a summative evaluation can only be performed on an object that is durable (i.e., capable of being presented time after time in substantially the same form) and reliable (i.e., capable of producing substantially the same results across many replications). These conditions were not satisfied in the case in point.

If this example were an isolated case, it would hardly justify comment, but it is not. Within the past year and one-half, I have encountered at least two similar situations in the field of health education. Both times, the requests for the "summative evaluation" have come from individuals who, while perhaps not laying claim to the title, effectively function as instructional developers. Diffusion of the concepts of formative and summative evaluation

seems to have had a concomitant of blurring of the concepts. In the diffusion process, there seems to have been introduced a certain laxness in definition of terminology. I propose that the terms "formative evaluation" and summative evaluation" have been adequately defined by Scriven, and are, in themselves, perfectly serviceable terms. Abuse of the terminology, however, has led to imprecision in communication, and I think it behooves us to rectify the problem by ensuring that the terms, and others like them, are applied only in those situations where they fit. If it arises that a certain kind of evaluation needs to be discussed, and it hasn't a name, let us give it a name rather than trying to extend an existing term beyond its scope.

Dichotomy or Trichotomy of Evaluation Roles?

In instructional development projects, situations occur which are neither fish nor fowl. For example, imagine that a curriculum developed elsewhere has been adopted, and some time after its implementation, it is discovered through a process of evaluation that the curriculum must be modified somewhat to meet peculiarly local needs. Is the evaluation that might turn up such a finding a formative evaluation, or a summative evaluation?

I am currently preparing a paper (Misanchuk, Note 1) in which I argue that the formative-summative description set ought to be expanded to include a third element, confirmative evaluation, which would describe that kind of an evaluation that is done on a product of an instructional development effort (be it a curriculum, a program, a set of learning materials, or whatever) that has been put into practice for a period of time, and is now up for review. That is, the product, which has been the subject of formative evaluations and possible of a summative evaluation, has been in operation for some time and the question now to be answered is "Should the product be maintained as is, changed in some way, or discarded completely, with or without replacement?"

Confirmative evaluation is sufficiently different than either formative evaluation or summative evaluation to warrant a designation of its own. First, consider the matter of timing. Formative evaluation is performed while the product is under development (i.e., before it is finished); summative evaluation is performed after the product is finished, but before it is widely adopted and implemented; confirmative evaluation is performed after the product has been implemented and used for a period of time. Second, consider the question of who is most likely to be able to perform a credible confirmative evaluation. Formative evaluation is best performed by an evaluator who is intimately involved in the development process; summative evaluation is best performed by an evaluator who is sufficiently removed from the project to be able to be objective and unbiased; confirmative evaluation requires something of both evaluator roles. To make the first decision of the three-stage hierarchy (see Figure 1) requires the dispassion of a summative evaluator, but to make the next two decisions requires the involvement of a formative evaluator. Perhaps a team effort is required for proper confirmative evaluation. Only further experience with conducting confirmative evaluations will produce the need guidelines. In any event, confirmative evaluation is clearly different than either formative or summative evaluation, both in its role and, consequently, in its demands.

The problem of precision in communication about evaluation in instructional development does not end with cleaning up the use of the formative-summative distinction. I submit that there is need for an expanded vocabulary of evaluation, especially in instructional development contexts. Let me illustrate that need with a couple of examples of evaluations in ID projects.

Example Case No. 1

Consider the evaluation of a human geography course (Schwen & Keller, 1977; Schwen, Keller, Backler & Jones, 1974) which

involved, among other things, measures of student achievement, lesson by lesson accumulation of student opinions with respect to several course components, and what Schwen and his colleagues termed a "follow-up evaluation" (1977, p. 33) which attempted to evaluate whether students were able to "think like geographers." The achievement tests were of the classic type, and the opinions on course components were collected from the 150 students primarily by Likert-type questionnaires. The data collected were therefore numerical in nature. The object of the evaluation, the human geography course, while relatively innovative in its format, with auto-tutorial laboratory, large group sessions, and discussion sessions, was pretty much a known quantity in the sense that university courses have long and often been the subject of evaluative scrutiny. For the most part, the data collected dealt with how well the students had learned the material presented in the course. A final assignment on the Applications Phase of the course, in which students were asked to use raw data and journal articles to demonstrate their ability to apply the concepts, principles, and skills taught in the course, was also present.

Example Case No. 2

In the second instance, consider the type of formative evaluation most of us practice when developing new products involving innovative formats or structures (e.g., see Baker, 1974, 1977; Markle, 1967): the evaluation consists largely of placing a single student into the instructional situation and observing him proceed through the prescribed learning sequence, then discussing his experiences, noting areas of difficulty or confusion. Although the process might be repeated with two or three students, it is generally performed with one learner at a time. The data collected are descriptions of specific points or generalizations made by the students, and non-participant

observations made by the developer/evaluator. The form of the evaluation is relatively familiar to instructional developers in the sense that most of us have done this procedure many times and have developed a more or less standard approach to the experience that may well vary somewhat in response to the particular situation, but is generally of similar format from instance to instance. Most frequently, the data collected will not be recorded in formal form (often they are simply notes in the margin of the prototype); seldom are the data numerical. Typically, the data collected deal with how quickly and easily the students are able to master the material presented (i.e., some sort of more or less formal criterion test on the taught material forms part of the data collection).

Example Case No. 3

Now consider still another example of a formative evaluation, in which I was involved recently. The evaluation problem had to do with determining whether the process by which development personnel were arranged and managed on a particular project was an effective one. The evaluation was formative, in the sense that the outcomes of the evaluation would be used to guide future decision-making on the pattern of ongoing deployment of human resources in the development project; it was also quite unique in the experience of the evaluator and the subjects of the evaluation. While there is a rich pool of information with respect to using pencil and paper instruments to collect evaluation information in numerical form and analyze such data, and even at least minimal suggestions on how to conduct the single-student feedback evaluations described in the second example, there seemed nowhere to turn for specific help in designing and implementing the type of evaluation described here. The situation was relatively novel, and there was little precedent available to draw upon. In terms of the focus of the evaluation, the behavior of the principal actors in the process was the object of scrutiny; it was not enough to know that each

participant in the development project knew what the development and management models in use were and how they operated--they had to abide by them in practise in order to judge the trial successful.

The type of evaluation described in the three examples is all formative, yet the nature and scope of the evaluations varied widely. Simply to describe all three evaluations as formative evaluations leaves a great deal undescribed. With an enlarged vocabulary pertaining to evaluation, we could describe the evaluations much more fully and richly, while retaining verbal parsimony.

Descriptors for Evaluations

For example, Davies and Schwen (1972) suggested the terms heuristic and algorithmic, and determinate and indeterminate to describe evaluations. The first of these two descriptive continuua refers to the form, or the how of the evaluation. Evaluation is heuristic if it is novel, unstructured, or consequential and involves no standardized, conventional, or agreed-upon approach. Algorithmic evaluation involves the application of established or standardized procedures, with little opportunity for deviation. The second set of descriptors, determinate and indeterminate, refers to the kind of data used to guide decision-making in the course of evaluation. Determinate evaluation decisions are typically based on numbers, are objective, analytical, and frequently computational in nature. In other words they use so-called "hard" data. Indeterminate decisions are typically based on the more subjective sources of intuition and experience (i.e., so-called "soft" data).

In the paper in preparation I mentioned earlier, I am suggesting that the substance, or the what, of the evaluation can be described in terms of acquirement evaluation and accomplishment evaluation. The former term refers to situations where the

evaluation is designed to determine the amount of learning that has occurred; the latter term refers to evaluations which not only determine the amount of learning that has occurred, but goes on to determine also whether or not he learned knowledge has sufficient social utility to the learner that he has put it into practice.

Using these three sets of terms allows us to differentiate rather better among the three examples of formative evaluation given earlier. The evaluation of the instructional development process that formed the basis of the third example could be described as heuristic (the situation was novel and unstructured, and suggested no standardized, conventional, or agreed-upon approach to the evaluation), indeterminate (decisions would be based on subjective data), and focussed on accomplishment (the integrated practice of the process was the focus of the evaluation, not just the knowledge of the process). The new product evaluation described in the second example could be characterized as algorithmic (most of the process follows an established procedure), indeterminate (the judgements are based on non-numerical, relatively subjective data), and focussed primarily on acquirement (i.e., how well the student learned the material presented). The evaluation of the human geography course in the first example was algorithmic (data collection was very structured), determinate (almost all the data involved were numerical), and acquirement-oriented (i.e., primarily focussed on the learing of the subject matter).

(The Applications Phase assignment in the human geography course evaluation attempted to get at a demonstration of utilization of the knowledge, skills, and concepts taught in the course, but the very fact that the demonstration was required under the circumstances of a formal assignment move the evaluation from the category of accomplishment to that of acquirement.

Accomplishment evaluation can only take place when there are no inducements to demonstrate use of the knowledge except the availability of the knowledge and the judgement of its applicability and efficacy.)

The Necessity for Descriptors

While it may be that, say, the new product evaluation process in the second example is less algorithmic than the human geography evaluation process (i.e., there is more opportunity for and more likelihood of variation in the data-collection process), it is at least meaningful to speak of the characteristics of the two evaluations in relative terms, by referring to their positions on the algorithmic-heuristic continuum.

There will be those who argue that what the field of education needs is less jargon, not more. The argument is that considerable obfuscation occurs as the result of using jargon. I must admit that when I encounter examples of poorly-constructed communications (including those of my own creation), I am sympathetic to the point. However, it seems that I far more frequently encounter situations in which terms are misused or inappropriately applied for want of an appropriate term. To me, there seem only two solutions to the problem--either a fuller description must be provided, using lay terms, to give an accurate description; or parsimony may be maintained by defining terms with specific meanings to fill the gap. It is in the spirit of the latter solution that I have written this paper.

The terms formative evaluation and summative evaluation are very useful for describing the roles of evaluations. By themselves, however, they are insufficient, to the extent that there are other characteristics of evaluations that ought to be described for inter-professional communications. Indeed, the formative-summative distinction does not cover all the available ground in the context

of instructional development, thus I have suggested that the term confirmative evaluation be added to the original dichotomy. Similarly, I have here reiterated the Davies-Schwen suggestion that the terms heuristic or algorithmic, and determinate and indeterminate be used to describe evaluations, and have suggested that the terms accomplishment and acquirement also be used.

A complete description of an evaluation, it seems to me, ought to include answers to the same basic questions that a good news story does. Most descriptions of evaluations do explicitly and implicitly inform the observer who is doing the evaluation, and where it is taking place. Use of the formative-summative-confirmative distinction adds information regarding the why of the evaluation, and also, incidentally, the when, since formative evaluation takes place during development, summative evaluation after development, but before implementation, and confirmative evaluation after implementation. The heuristic-algorithmic dimension attends, as does the determinate-indeterminate dimension, to the how of the evaluation, by providing detail regarding the procedures and data types used in the evaluation. Finally the acquirement-accomplishment dimension addresses itself to the question of what kind of outcome or behavior is under scrutiny.

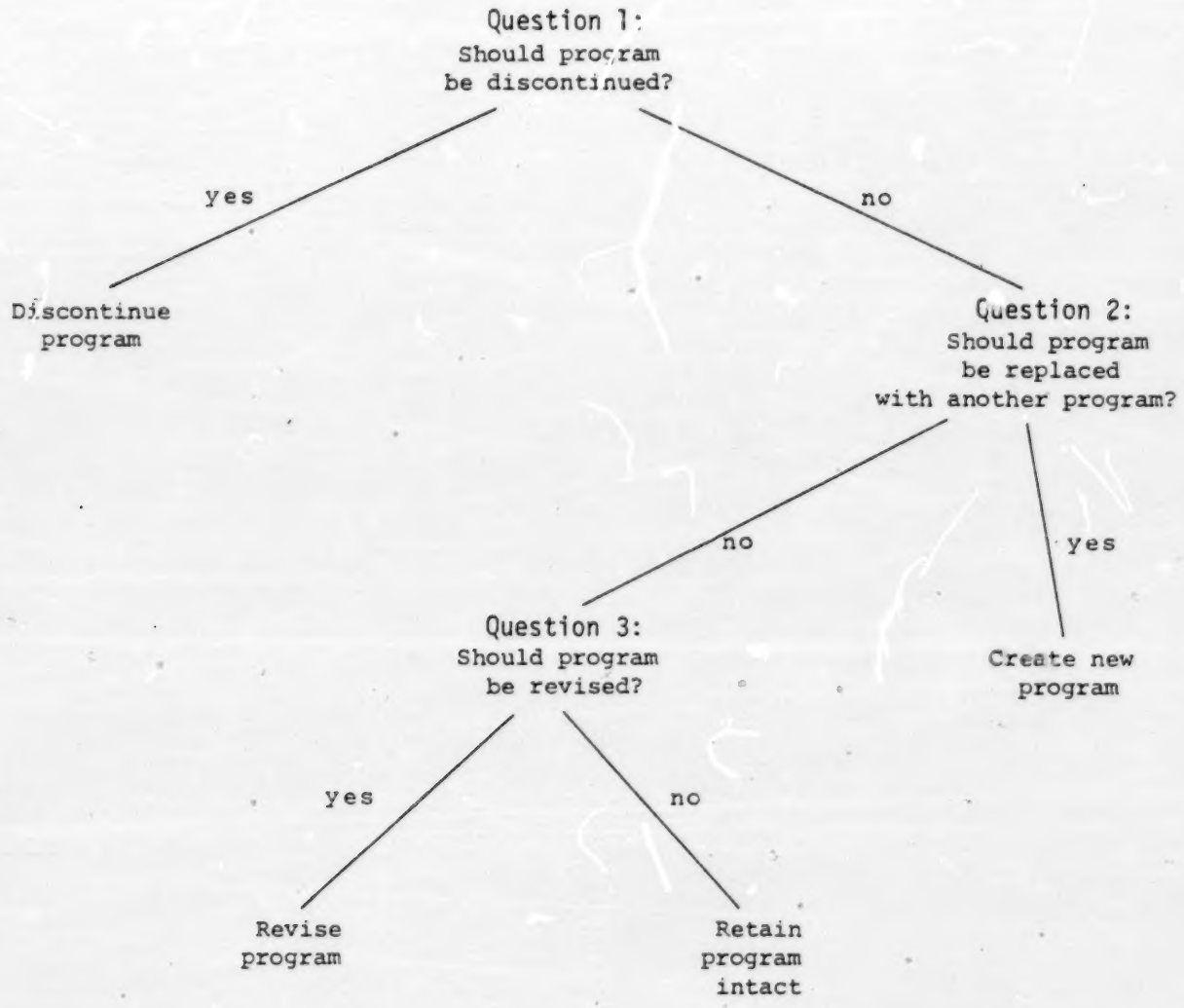

Figure 1: Decision stages in confirmative evaluation. Question 1 resembles a summative evaluation question; Question 3 resembles a formative evaluation question; Question 2 has aspects of both formative and summative evaluation questions.

REFERENCE NOTES

Misanchuk, E.R. Descriptors of evaluation: The formative-summative distinction in evaluation and some additional dimensions. Paper in preparation.

REFERENCES

Baker, E.L. (1974) Formative evaluation of instruction. In W.J. Popham (Ed.), Evaluation in education: Current applications. Berkeley, Calif.: McCutchan.

Baker, E.L. (1977) The dance of evaluation: Hustle or minuet? Journal of Instructional Development, 1(1), 26-28.

Davies, I.K., & Schwen, T.M. (1972) Some process concerns about formative evaluation. Viewpoints, 48(4), 1-8.

Markle, S.M. (1967) Empirical testing of programs. In P.C. Lange (Ed.), Programed instruction: The 66th yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press.

Schwen, T.M., & Keller, J.M. (1977) A case study: Developing convergent formative evaluation methodology. Journal of Instructional Development, 1(1), 31-35.

Schwen, T.M., Keller, J.M., Backler, A.L., & Jones, D.L. (1974) Formative evaluation of the development of a human geography course. Bloomington, Ind,: Indiana University, Audio Visual Center, Division of Development and Special Projects. Evaluation Document No. 1.

Scriven, N. (1967) The methodology of evaluation. In R.W. Tyler (Ed.), Perspectives of curriculum evaluation. Chicago: Rand McNally.